Detecting individual change in children with mild to moderate motor impairment: The Standard Error of Measurement of the Movement ABC

Leemrijse, C.; Meijer, O.G.; Vermeer, A.

published in
Clinical Rehabilitation
1999

DOI (link to publisher)
10.1191/026921599675491988

Clinical Rehabilitation 1999; **13**: 420–429

Detecting individual change in children with mild to moderate motor impairment: the standard error of measurement of the Movement ABC

Chantal Leemrijse Department of Occupational Therapy, **Onno G Meijer** Faculty of Human Movement Sciences, Vrije Universiteit, Amsterdam, **Adri Vermeer** Faculty of Social Sciences, Utrecht University, Utrecht, **Brigitte Lambregts** Department of Occupational Therapy, Hospital De Heel, Zaandam and **Herman J Adèr** Department for Epidemiology and Biostatistics, Vrije Universiteit, Amsterdam, The Netherlands

Received 20th August 1998; returned for revisions 18th September 1998; revised manuscript accepted 20th December 1998.

Objective: To assess whether the Movement ABC can be used to monitor individual change in motor performance.
Design: Motor-impaired children were tested three times in succession with the Movement ABC without any intervention.
Setting: Two schools for special education and one school for children who are chronically ill.
Subjects: Three girls and 20 boys aged 6–8 years.
Main outcome measures: Scores were measured per item (0 → 5), added to cluster scores (0 → 10 or 15), added to form the total scores (0 → 40). Mean scores, standard errors of measurement (SEMs) and least detectable differences (LDDs) were calculated per item, per cluster and for the total scores. A repeated measures analysis of variance was performed to test for the effects of time.
Results: The total scores improved significantly from the first session (mean: 15.4 points) to the second (mean: 13.3), but not from the second to the third (mean: 13.2). Average item scores ranged from 0.6 to 2.7 points with SEMs of 0.79 → 1.54 and LDDs of 2.20 → 4.27. Average cluster scores ranged from 3.4 to 5.3 with SEMs of 1.51 → 1.84 and LDDs of 4.18 → 5.11. The SEM of the total scores equalled 3.13 with an LDD of 8.68.
Conclusions: The total score of the Movement ABC is sufficiently sensitive to monitor individual change; the cluster scores have moderate sensitivity and individual items are inappropriate to monitor individual change. The significant effect of time is interpreted as an effect of learning.

Address for correspondence: Chantal Leemrijse, Department of Occupational Therapy, VU Hospital, De Boelelaan 1108, 1007 MB Amsterdam, The Netherlands. e-mail: C.Leemrijse@azvu.nl

 0269–2155(99)CR255OA

Introduction

Standardized tests play an important role in occupational and physical therapy for children with developmental delays. While basic information concerning each individual child will always come from parents and teachers, and from medical records, paediatric therapists have their own specific contribution to the diagnostic and evaluative process. The clinical observations of a paediatric therapist allow for a professional assessment of the functional limitations of the individual child.[1,2] Standardized paediatric tests were developed to give these observations an objective basis. Standardized tests are crucial not only to the diagnostic process but also to the planning and evaluation of treatment.

A widely used norm-referenced test to evaluate motor functioning in children is the Movement ABC. The Movement ABC was developed to serve two main purposes. First, the test can have a diagnostic function in offering a detailed picture of the child's motor weaknesses and strengths. Second, the movement ABC is reported to be sufficiently reliable and valid to use in evaluating the efficacy of treatments for motor impaired children.[3] Indeed, several intervention studies[4–7] reported significant effects in groups of motor-impaired children who were tested with the Movement ABC or its predecessor, the Test of Motor Impairment – Henderson's revision.[8] Without denying the importance of the published group studies, we want to emphasize that in clinical practice the paediatric therapist focuses on the individual child. In the literature, information is lacking as to the usefulness of the Movement ABC in monitoring individual change in motor-impaired children. The purpose of this paper is to start obtaining such information.

Henderson and Sugden[3] reported a test–retest agreement from 62% to 100% for the items of the Movement ABC, when scores were dichotomized into pass or fail. Such stability or instability of test scores is usually interpreted as a quality of the measurement instrument, while in fact the subjects may have changed between test and retest. Although research tends to neglect such fluctuations, it is the common experience of paediatric therapists that the degree to which a child suffers from functional limitations may vary from day to day.[9] Hence it is often difficult to differentiate the variance that is caused by the measurement procedure (including the person who takes the measurements) from that which is intrinsic to the measured property itself. In our opinion, therapists just transcend this problem by focusing on the clinical distinction between meaningful and trivial change. In statistical terms meaningful change at least requires systematic variance rather than unsystematic variance. Examples of sources of systematic variance are maturation of the child, specific learning effects on the test items, or the effects of treatment. Unsystematic variance consists of transient fluctuations in the behaviour of the child plus the classical unreliability of the test instrument itself.

A useful statistic that presents itself for differentiating between systematic change and unsystematic variance is the standard error of measurement (SEM). The SEM provides an estimate of the unsystematic variance in a subject's score if a test is administered more than once under the same conditions.[10–19] Since children with motor impairment do not form a homogeneous group,[20–23] one would theoretically want to establish a separate SEM for each individual child by measuring the same child many times. Obviously, this would be very impractical, if not impossible. Therefore, the SEM is estimated by measuring a group of comparable subjects a few times.

In the present study, a group of motor-impaired children were measured three times with the Movement ABC and the SEM was estimated. Based on the SEM, a 'least detectable difference' (LDD) was derived, representing the minimal individual change that must be observed before the therapist may conclude that systematic change has occurred.

Methods

Movement ABC and procedure

The Movement ABC was specifically developed for children with motor impairment without mental retardation or any known physical disorder such as cerebral palsy. To date, such children are said to suffer from 'developmental coordination disorder' (DCD).[24] DCD is found about three times more often in boys than in girls.[25] In the literature, children scoring at or below the 5-centile of the Movement ABC are regarded as children with 'definite' motor problems.[3] In the present study, we defined as 'mild to moderate motor impairment' scores below the 30-centile, that is, scores of children who do experience themselves to be clumsy in comparison to many of their peers without necessarily having a definite motor problem.

The Movement ABC has two parts: a battery of eight different performance tests for the individual child, and a checklist to be filled in by an adult who is familiar with the child's day-to-day motor functioning.[3] The checklist is more informal and is not taken into account in the scores on the performance tests. The present study focused on the performance tests only. While these tests are slightly different per age-band, the tasks to be performed are very comparable.[3] Score distributions and variability of scores are similar for different age groups (6–12 years), and gender differences are small and nonsignificant.[3] The present study was performed with children from 6 to 8 years old, and only the first two age-bands of the performance tests had to be used (Table 1).

Items are scored in seconds or number of correct trials. The manual gives a strict protocol to translate these raw scores into point scores (between 0 and 5). For most items only integers are used but for items 1 and 6, half points can be given also. For all items, higher scores indicate a more serious level of impairment. The point scores for the individual items are added to form three clusters: manual dexterity (items 1–3), ball skills (items 4 and 5) and balance (6–8). Finally, the cluster scores are added to form the total scores. It may be important to note that the item scores are at least ordinal but not necessarily interval, while the total scores can be treated as interval data.[3] A large sample of norm-reference children has been studied,[3] which allows item, cluster and total scores of the Movement ABC to be translated into centiles (Table 2).

In our study, the Movement ABC was used in

Table 2 Centile equivalents of point scores on the Movement ABC

	Points	Centile
Items	0.0	≥75
	1.0	10–75
	2.0–3.0	5–10
	4.0	3–5
	5.0	<2
Manual dexterity	5.0	<15
	6.5	<5
Ball skills	2.5	<15
	5.0	<5
Balance	5.0	<15
	7.5	<5
Total[a]	7.0	<30
	10.0	<15
	13.5	<5

[a]For the total score, more detailed centile equivalents are available.[3]

Table 1 Items of the Movement ABC (performance tests); first two age-bands

	Band 1: 4, 5 and 6 years old	Band 2: 7 and 8 years old
Manual dexterity		
Item 1	Posting coins	Placing pegs
Item 2	Threading beads	Threading lace
Item 3	Bicycle trail	Flower trail
Ball skills		
Item 4	Catching bean bag	One hand and bounce
Item 5	Rolling ball into goal	Throwing bean bag into box
Balance		
Item 6	One leg balance	Stork balance
Item 7	Jumping over cord	Jumping in squares
Item 8	Walking heels raised	Heel to toe walking

accordance with the directions specified in the manual.[3] The measurements were performed by two experienced occupational therapists who had been specifically trained in the administration of the Movement ABC. Each child was tested by the same examiner all three times. Time intervals were planned that were long enough to minimize temporary practice effects and short enough to render systematic effects of maturation implausible. For practical reasons, the planned time interval between testing sessions varied unsystematically from two (31% of the cases) to three (65% of the cases) weeks. Two children could not make the planned interval. For one of them, the second session was four weeks after the first; for the other, the third session was four weeks after the second. Tests were performed on the same weekday and at the same time during school hours, unless illness of the child precluded this. The same quiet test room was used for all sessions.

Subjects

Children were selected from three primary schools in The Netherlands: two schools for children with learning disorders and one school for chronically ill children. In such schools for special education, there is usually a high percentage of children with motor difficulties. Teachers, and at one school the physical therapist of the school, were asked to select pupils between 6 and 8 years old who had poor motor co-ordination for their age, which interfered with their school performance. This age range was chosen because it is usually the range in which co-ordination problems are first 'officially' recognized, that is, by teachers or school medical staff.

Children with a marked physical disability (such as visual impairments or orthopaedic abnormalities) and/or known neurological disorders (such as cerebral palsy or epilepsy) were not considered for participation in the study. Furthermore, children had to master the Dutch language. Parents of those children who fulfilled the above criteria were asked to fill in an informed consent. All children for whom consent was obtained were then tested with the Movement ABC. Children who scored above the 30-centile at this first test session were excluded from the second and third sessions.

Statistical analysis

In order for an LDD (least detectable difference) to be meaningfully interpretable, the frequency distribution in question should be sufficiently close to normal. We operationalized this 'closeness' in terms of skewness and kurtosis. Skewness and kurtosis of all item scores, the cluster scores and total scores were determined, first per session, then for the three sessions taken together. It was decided to accept only variables with skewness as well as kurtosis between –3 and +3 in all cases. Note that some skewness had to be expected since some of the children were bound to do well in some of the sessions on some of the items. In order to see how strong this effect would be in our sample, we also calculated the percentage of zeros for each item, the cluster scores and the total scores.

For every test session, mean scores, SDs and score ranges were calculated per item, per cluster and of the total score. Analysis of variance (ANOVA) for repeated measures was performed on the total score, the cluster scores and each item separately. Such a repeated measures ANOVA allows for removing the between-subjects variance in order to focus on the within-subjects variance,[26] which is the variance of interest in the present study. This within-subjects variance is further partitioned into between-trials variance and residual variance.[10] The between-trials variance contains the systematic differences over time and can be used to pinpoint learning or maturation. In the present study, Helmert contrasts[26] were calculated for the first measurement versus the average of the second and the third, and for the second measurement versus the third.

The residual variance is what is left after removing both the between-subjects variance and the systematic effects of time.[26] This residual variance is used to calculate the SEM. It is important to note that the residual variance, and thus the SEM, combines unsystematic variance in the behaviour of the subjects with classical measurement errors, including the unsystematic contributions to the variance of both examiners. Thus, for the clinician who wants to monitor individual change, the SEM takes together all unsystematic factors one wants to contrast with 'real' change. The SEM is estimated by taking the square root of the residual variance[18]:

$$SEM = \sqrt{MS_{residual}}$$

Since the SEM is expressed in the same units as the measurement itself, it can be interpreted as a standard deviation.[10–19] If the scores of an individual can be assumed to have a Normal distribution, the 95% confidence interval of any single score X equals: $X \pm 1.96 \times SEM$. The SEM of the difference between two subsequent scores of the same individual is calculated as: $SEM_{difference} = \sqrt{2} \times SEM$. Taking 5% as the significance level, the minimum difference that allows one to conclude for a 'real', that is significant change, thus equals[11–18]:

$$1.96 \times \sqrt{2} \times SEM$$

This quantity, $1.96 \times \sqrt{2} \times SEM$, defines the least detectable difference (LDD) between two subsequent scores. If the actual difference between two subsequent scores exceeds the LDD, statistically significant change has occurred ($p < 0.05$). With respect to our focusing on the LDD, it may be important to note that the notions of discriminative and evaluative validity are different in principle. In order for the Movement ABC to be valid in diagnosis, it should validly discriminate between children. This is why the Movement ABC is a norm-referenced test.[3] On the other hand, the evaluative decision if an individual child has 'really' improved, depends on the choice of a criterion. In the present study, we focused on the criterion that the change should be statistically significant, i.e. exceed the LDD.

The LDD was computed for each item, each cluster, and the total score. In terms of the LDD, children can improve significantly between two subsequent scores only if their initial score exceeds one LDD. We determined per item, for the clusters and for the total scores how many of the children started with scores below one LDD and for whom improvement by at least one LDD would thus have been impossible. Furthermore, if the score range is only slightly larger than the LDD, significant improvement between subsequent scores is not possible more than once. We therefore also expressed the LDD as fraction of the score range. Finally, in the Discussion we give a more clinical interpretation of the LDD.

Throughout the analysis, SPSS 7.5 for Windows was used. All statistical tests were performed with a significance level of 0.05.

Results

Parental consent was obtained for 28 children. Five of them had initial scores on the Movement ABC above the 30-centile. These children were removed from the study. Of the remaining 23 subjects (Table 3), 18 scored below the 15-centile, 13 of them below the 5-centile. Thus, 13 of the 23 children in the present study had 'definite' motor problems.[3] There were 20 boys and 3 girls in our sample. This 20:3 ratio exceeds the 3:1 ratio given in the literature.[25] Since the Movement ABC is reported to be robust for gender differences,[3] this over-representation of boys was not treated as a confounding factor. There was no experimental attrition.

For every test session, mean scores, SDs and score ranges were calculated (Table 4). Note that the whole range of possible scores was covered 19 out of 24 times in the item scores, 2 out of 9 times in the clusters (both in ball skills), but never in the total scores. In item 7, the kurtosis was larger than +3 in the first session and in the total distribution of scores. Thus, the LDD of item 7 is not meaningfully interpretable. For item 7, 81% of all scores equalled 0. On average, 36% of all item scores equalled 0, 10% of cluster

Table 3 Subject characteristics

	Age		
	6 years	7 years	8 years
Number of subjects	7	7	9
Ratio male/female	0/7	1/6	2/7
Total score first measurement:			
Mean	16.5	15.2	14.8
SD	6.1	3.4	7.7

Table 4 Mean scores, SDs and score ranges per session

	First session			Second session			Third session		
	Mean	SD	Range	Mean	SD	Range	Mean	SD	Range
Item 1	1.5	1.4	0.0 → 4.5	1.3	1.6	0.0 → 5.0	1.5	1.3	0.0 → 4.0
Item 2	2.6	1.8	0.0 → 5.0	2.0	1.6	0.0 → 4.0	2.1	1.6	0.0 → 4.0
Item 3	1.2	1.4	0.0 → 5.0	1.4	1.6	0.0 → 5.0	1.2	1.4	0.0 → 5.0
Item 4	2.7	1.5	0.0 → 5.0	1.7	1.4	0.0 → 5.0	1.8	1.5	0.0 → 5.0
Item 5	2.7	1.7	0.0 → 5.0	1.7	1.6	0.0 → 5.0	2.1	1.5	0.0 → 5.0
Item 6	2.6	1.4	0.0 → 4.5	2.2	1.3	0.0 → 5.0	2.1	1.8	0.0 → 5.0
Item 7	0.6	1.2	0.0 → 5.0	0.7	1.6	0.0 → 5.0	0.7	1.7	0.0 → 5.0
Item 8	1.7	2.0	0.0 → 5.0	2.3	2.1	0.0 → 5.0	1.8	2.2	0.0 → 5.0
Manual dexterity	5.3	3.3	0.0 → 12.5	4.7	3.5	0.0 → 14.0	4.8	3.4	0.0 → 13.0
Ball skills	5.3	1.9	2.0 → 9.0	3.4	2.4	0.0 → 10.0	3.8	2.0	0.0 → 8.0
Balance	4.8	3.4	1.0 → 11.5	5.3	4.4	0.0 → 15.0	4.6	4.2	0.0 → 14.5
Total score	15.4	3.9	7.5 → 30.5	13.3	7.3	0.0 → 30.0	13.2	6.9	2.0 → 26.0

scores and 1% of the total scores. Apart from item 7, skewness and kurtosis were always between –3 and +3.

Repeated measures ANOVA (Table 5) revealed a significant effect of time on the total scores ($F_{2,44} = 3.62$; $p = 0.035$), the cluster score for ball skills ($F_{2,44} = 8.09$; $p = 0.001$) and the scores on item 4 ($F_{2,44} = 6.59$; $p = 0.003$). On Helmert contrasts, the average scores of the second and third session were significantly better than those of the first session (total scores $F_{1,22} = 6.49$, $p = 0.041$; ball skills $F_{1,22} = 13.73$, $p = 0.001$; item 4 $F_{1,22} = 10.96$, $p = 0.003$). There were no significant differences between the scores of the second and those of the third session.

The SEM was calculated as the square root of the residual variance in the above repeated measures ANOVA (Table 5). For the item scores SEMs ranged from 0.79 to 1.54, for the cluster scores from 1.51 to 1.84. The SEM for the total score equalled 3.13. The LDDs (Table 5) were

Table 5 Mean squares, SEMs and LDDs

	Mean square			SEM	LDD	
	Between subjects[a] (df = 22)	Between sessions (df = 2)	Residual (df = 44)			Score range (points)
Item 1	5.02	0.26	0.63	0.79	2.19	5
Item 2	6.19	2.62	0.97	0.99	2.74	5
Item 3	4.89	0.36	0.88	0.94	2.61	5
Item 4	4.64	6.57*	1.00	1.00	2.77	5
Item 5	3.15	5.84	2.37	1.54	4.27	5
Item 6	6.56	1.48	0.94	0.97	2.69	5
Item 7	4.88	0.19	1.17	1.08	2.99[b]	5
Item 8	10.35	2.88	1.41	1.19	3.30	5
Manual dexterity	30.45	2.31	2.28	1.51	4.19	15
Ball skills	7.65	24.10*	2.98	1.73	4.80	10
Balance	41.76	2.87	3.39	1.84	5.10	15
Total score	117.41	35.55*	9.82	3.13	8.68	40

[a]For all items, clusters and the total score, $p < 0.05$.
[b]Not meaningfully interpretable.
*$p < 0.05$.

calculated from the SEMs (cf. Statistical analysis). LDDs for the items ranged from 2.19 to 4.27, for the clusters from 4.19 to 5.10. The LDD of the total score equalled 8.68. Of the initial item scores in our sample, 86% fell below one LDD. This also occurred in 46% of the initial cluster scores and in 13% (three children) of the initial total scores. In all these cases, significant improvement of the individual child would, per definition, be impossible. Expressed as a fraction of the whole score range, LDDs of the item scores ranged from 0.44 (item 1) to 0.85 (item 5), LDDs of the cluster scores from 0.28 (manual dexterity) to 0.48 (ball skills). The LDD of the total scores equalled 0.22 of the total score range, showing that the total score is potentially much more responsive than the cluster scores or the item scores.

Discussion

In the present sample of children with mild to moderate motor impairment, we found an SEM of 3.13 points for the total score of the Movement ABC (scale range $0 \rightarrow 40$). Of course, it has to be realized that the number of children in our study was relatively small, and the generalizability of our results *per se* is limited.

In 1990, Riggen *et al.*[27] reported an SEM of 0.86 points in healthy preschoolers measured with the predecessor of the Movement ABC, the TOMI-H (scale range $0 \rightarrow 16$). Expressed as percentage of the total scale range, the SEM in our study is somewhat larger (8%) than that in the study of Riggen *et al.* (5%). A plausible explanation of this difference is that the performance of children with mild to moderate motor impairment is more variable than that of unimpaired children of the same age. It has to be noted that the scoring possibilities of the TOMI-H per item ($0 \rightarrow 2$ points) are fewer than those of the Movement ABC ($0 \rightarrow 5$ points). Thus, the TOMI-H is probably less sensitive to differences in motor performance.

In order to be sufficiently sensitive to monitor individual change, the Movement ABC should have LDDs that are much smaller than the score range. For the total score, this is clearly the case: an LDD of 8.68 points within a range of $0 \rightarrow 40$. In the cluster scores, the LDDs form a larger fraction of the score range, but still remain below 50%. Thus, the cluster scores appear to be moderately sensitive to individual change. At the level of the individual items, the situation is very different. The LDDs of the individual items cover 44% to 85% of the score range. The therapists in the present study were specifically trained to use the Movement ABC and large measurement errors were thus not to be expected. Hence, the only plausible explanation of the relative insensitivity of the individual items is that the actual performance of the children varies strongly from measurement to measurement at the level of the items.

We recalculated our results for the 18 children whose initial total scores were below the 15-centile. The general pattern remained the same. Even children below the 15th percentile often score very well on some individual items. The essence of the Movement ABC is to take item scores together. A child who now fails on this and then on another item, still can have stable scores at the level of the clusters and that of the total scores.

Given the above, it turns out to be undesirable to use the individual items of the Movement ABC to monitor individual change in children with mild to moderate motor impairment. Moreover, most initial item scores in our sample fell below one LDD, rendering significant improvement theoretically impossible. This again shows that the variability at the item level precludes the use of single scores in monitoring individual change. Nevertheless, the item scores can still be useful in comparing averages. In our sample, significant between-subject differences were found for all averages over the three sessions, even at the level of the individual items. Schoemaker *et al.*[5] reported significant differences between an experimental and a control group not only in the total scores but also in items 2, 4 and 6. And Smits-Engelsman *et al.*[6] found significant differences between groups in the balance score and in items 7 and 8 of the Movement ABC.

Clinical relevance

For the paediatric therapist in daily clinic, LDDs allow for a straightforward statistical interpretation of change scores: An individual child has significantly improved (p <0.05) between two consecutive measurements if the child has improved by at least one LDD. However, the clinical interpretation of SEMs and LDDs is far from straightforward. No single score makes it possible to differentiate measurement error from actual performance. In our study, significant improvement was found between the first and the second measurement session. Since there was only a short time interval between measurements and no intervention was given, these differences were interpreted as a learning effect. If a child is to be monitored with the Movement ABC, we recommend measuring the child twice at the beginning of therapy and discarding the first measurement.

If the above precaution is taken, however, there remains the problem of relating the statistical interpretation of change scores to their clinical interpretation. In group studies, it is possible to find significant differences that are too small to be clinically meaningful. We do not believe that the same problem exists if an individual child improves one LDD or more on the Movement ABC. Such an improvement of 9 points in the total score would be larger than from the 1-centile (18.5 points) to the 15-centile (10 points).[3] Children with mild to moderate motor impairment suffer from the fact that they are less competent in motor skills than their peers.[28] Although there is no absolute standard for clinical relevance, we would not doubt the meaningfulness of an improvement from the first to the 15th percentile. Accordingly, we propose to regard a significant improvement on the Movement ABC as a clinically meaningful improvement.[29]

On the other hand, assessing the clinical relevance of nonsignificant improvement is notoriously difficult. The LDD is a conservative measure and rather large improvement is needed for statistical significance. Thus, nonsignificant improvement is likely to occur frequently. If an individual child were to improve, for instance, by just over one SEM, 3.5 points on the total score, the change would be equal to that from the 5-centile to the 15-centile. Such nonsignificant improvement thus has the potential to be clinically meaningful. The challenge then becomes to decide if the improvement in question is of a systematic nature, that is, not just deriving from random fluctuations in the child's behaviour or measurement error. Skilled therapists make such decisions every day, often without consciously using any formal quantitative models.

No therapist ever relies on one source of information only. Parents and teachers continue to provide information; the therapist observes the child and interacts with him or her, not relying on single measures but observing the overall development over time; colleagues offer relevant information; there may be new results from medical tests, etc. In order to differentiate meaningful from trivial change, paediatric therapists take many sources of information into account. It is a well-known statistical fact that information from several independent sources taken together is more reliable than information from single sources.[30] Accordingly, we believe that skilled therapists regard potentially meaningful nonsignificant improvement as 'real' improvement when changes in other dependent variables point in the same direction.

Conclusion

Clinicians are increasingly being expected to use reliable, standardized and validated measures in their daily practice. Monitoring individual children with such measures may encourage the clinician to base their interventions on objective change. Objective improvement can offer a motive to continue or even intensify the intervention, while objective deterioration may lead to choosing a different therapeutic strategy. The present study shows that the total scores of the Movement ABC can be used to validly monitor change in children with mild to moderate motor impairment. The individual items of the Movement ABC are insufficiently sensitive for that purpose, while the cluster scores may offer some useful information but are less sensitive than the total scores.

Acknowledgements

Our special thanks go to the children, their parents, the teachers and the physical therapist at the Openluchtschool, the Jorisschool and the Schakel in Breda, as well as the Van Houtenschool in Amsterdam, without whom this study could not have taken place. We gratefully acknowledge Joan Verhoef for her academic contribution and her participation in the actual measurements. Joop Kuijk offered valuable statistical advice. Reint Geuze, Lex Kalverboer, Tim Koelman, Charles Njiokiktjien, Jacques Van Rossum and Robert Wagenaar gave useful commentary on an earlier version of this paper. Comments from the referees of *Clinical Rehabilitation* greatly helped to improve the readability of the paper; in fact, we shamelessly borrowed some of their phrases in our Discussion and Conclusion. Thanks. Sandra Diemel was very helpful when the paper was given its final shape.

The research was made possible by grants from the JANIVO foundation, the Fonds voor Muziektherapie of BUMA/STEMRA, and the Stichting Kinderpostzegels Nederland.

References

1 Jette AM. Physical disablement concepts for physical therapy research and practice. *Phys Ther* 1994; **74**: 379–86.
2 Van der Net JJ. Functional aspects of rheumatic diseases in childhood. Doctoral Dissertation, Utrecht University, The Netherlands, 1995.
3 Henderson SE, Sugden DA. *Movement Assessment Battery for Children*. Kent: Harcourt Brace Jovanovich, 1992.
4 Laszlo JI, Bairstow PJ, Bartrip J, Rolfe UT. Clumsiness or perceptuomotor dysfunction? In: Colley AM, Beech JR eds. *Cognition and action in skilled behaviour*. Amsterdam: Elsevier, 1988: 293–310.
5 Schoemaker MM, Hijlkema MGJ, Kalverboer AF. Physiotherapy for clumsy children: An evaluation study. *Dev Med Child Neurol* 1994; **36**: 143–55.
6 Smits-Engelsman BCM, Schoemaker MM, Jansen MPHT, Niemeijer AS. Physical therapy for children with writing difficulties. [In Dutch] *Ned Tijdschr Fysiother* 1996; **106**: 156–66.
7 Sims K, Henderson SE, Hulme C, Morton J. The remediation of clumsiness, 1: An evaluation of Laszlo's kinaesthetic approach. *Dev Med Child Neurol* 1996; **38**: 976–87.
8 Stott DH, Moyes FA, Henderson SE. *Test of Motor Impairment – Henderson revision*. San Antonio: Psychological Corporation, 1984.
9 Van Geert P. *Dynamic systems and development: change between complexity and chaos*. New York: Harvester Wheatsheaf, 1994.
10 Rothstein JM. *Measurement in physical therapy*. New York: Churchill Livingstone, 1985.
11 Guyatt G, Walter S, Norman G. Measuring change over time: assessing the usefulness of evaluative instruments. *J Chron Dis* 1987; **40**: 171–78.
12 Rogers JC. Selection of evaluation instruments. In: King-Thomas MHS, Hacker BJ eds. *A therapist's guide to pediatric assessment*. Boston: Little, Brown and Company, 1987: 19–33.
13 Deitz JC. Reliability. *Phys Occup Ther Pediatr* 1989; **1**: 125–47.
14 Derstine S. Tests of infant and child development. In: Tecklin JS ed. *Pediatric physical therapy*. Philadelphia: Lippincott, 1989: 16–39.
15 Stratford P. Reliability: Consistency or differentiating among subjects? *Phys Ther* 1989; **69**: 299–300.
16 Cunningham Amundson SJ, Crowe TK. Clinical applications of the standard error of measurement for occupational and physical therapists. *Phys Occup Ther Pediatr* 1993; **12**: 57–71.
17 Roebroeck ME, Harlaar J, Lankhorst GJ. The application of generalizability theory to reliability assessment: an illustration using isometric force measurements. *Phys Ther* 1993; **73**: 386–401.
18 Eliasziw M, Young SL, Woodbury MG, Fryday-Field K. Statistical methodology for the concurrent assessment of inter rater and intra rater reliability: using goniometric measurements as an example. *Phys Ther* 1994; **74**: 777–88.
19 Wilson BN, Polatajko HJ, Kaplan BJ, Faris P. Use of the Bruininks–Oseretsky test of motor proficiency in occupational therapy. *Am J Occup Ther* 1995; **49**: 8–17.
20 Hoare D. Subtypes of developmental coordination disorder. *Adapt Phys Act Q* 1994; **11**: 158–69.
21 Shanks Sellers J. Clumsiness: Review of causes, treatments and outlook. *Phys Occup Ther Pediatr* 1995; **15-4**: 39–55.
22 McConell D. Processes underlying clumsiness: a review of perspectives. *Phys Occup Ther Pediatr* 1995; **15-3**: 33–52.
23 Wright HC, Sugden DA. The nature of developmental coordination disorder: inter- and intragroup differences. *Adapt Phys Act Q* 1996; **13**: 357–71.
24 *DSM-IV diagnostic and statistical manual of mental disorders*, fourth edition. Washington: American Psychiatric Association, 1994.
25 Henderson SE, Hall D. Concomitants of clumsiness in young school children. *Dev Med Child Neurol* 1982; **24**: 448–60.
26 Stevens JP. *Intermediate statistics: a modern approach*. Hillsdale, NJ: Lawrence Erlbaum, 1990.
27 Riggen KJ, Ulrich DA, Ozmun JC. Reliability and concurrent validity of the Test of Motor Impairment – Henderson revision. *Adapt Phys Act Q* 1990; **7**: 249–58.
28 Schoemaker MM, Kalverboer AF. Social and

affective problems of children who are clumsy: how early do they begin? *Adapt Phys Act Q* 1994; **11**: 130–40.

29 Beckerman H, Roebroeck ME, Becher J, Lankhorst GJ, Bezemer PD, Verbeek ALM. The sickness impact profile: a useful outcome measure in stroke patients? Test–retest reproducibility and its implications for responsiveness. In: Beckerman H ed. Efficacy of thermocoagulation and an ankle-foot orthosis in stroke patients: A placebo controlled randomized clinical trial. Doctoral Dissertation, Vrije Universiteit Amsterdam, The Netherlands, 1996.

30 Fleiss JL. Reliability of measurement. In: Fleiss JL ed. *Design and analysis of clinical experiments*. New York: John Wiley, 1986: 1–32.